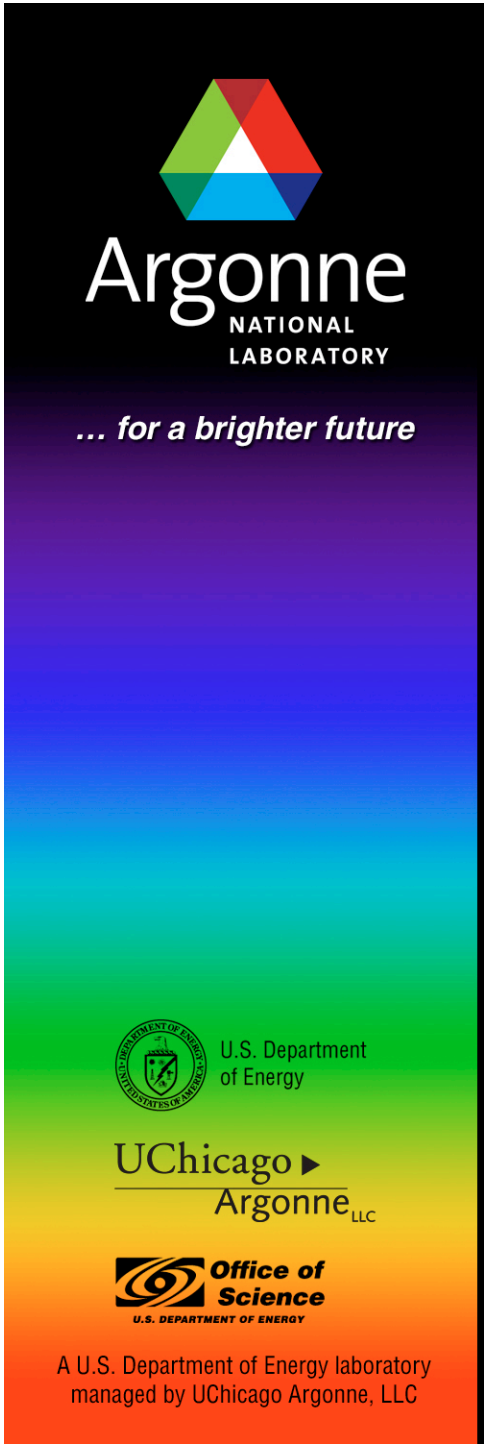


Vis Breakout

Mark Hereld & Joe Insley

ALCF's Eureka and Gadzooks

- 104 dual quad core servers
- 208 Quadro FX5600 graphics engines
- 312 Gbytes of total frame buffer RAM
- 3.2 TB of total system RAM

- Each node (server):
  - Dual quad core CPU
  - 2 GPUs
  - 1.5 GB frame buffer RAM
  - 32 GB system RAM

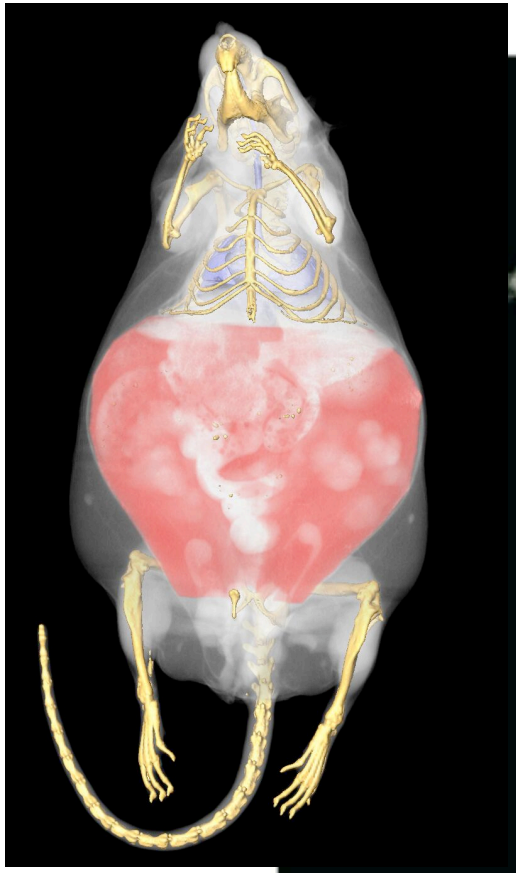
Under the Circumstances

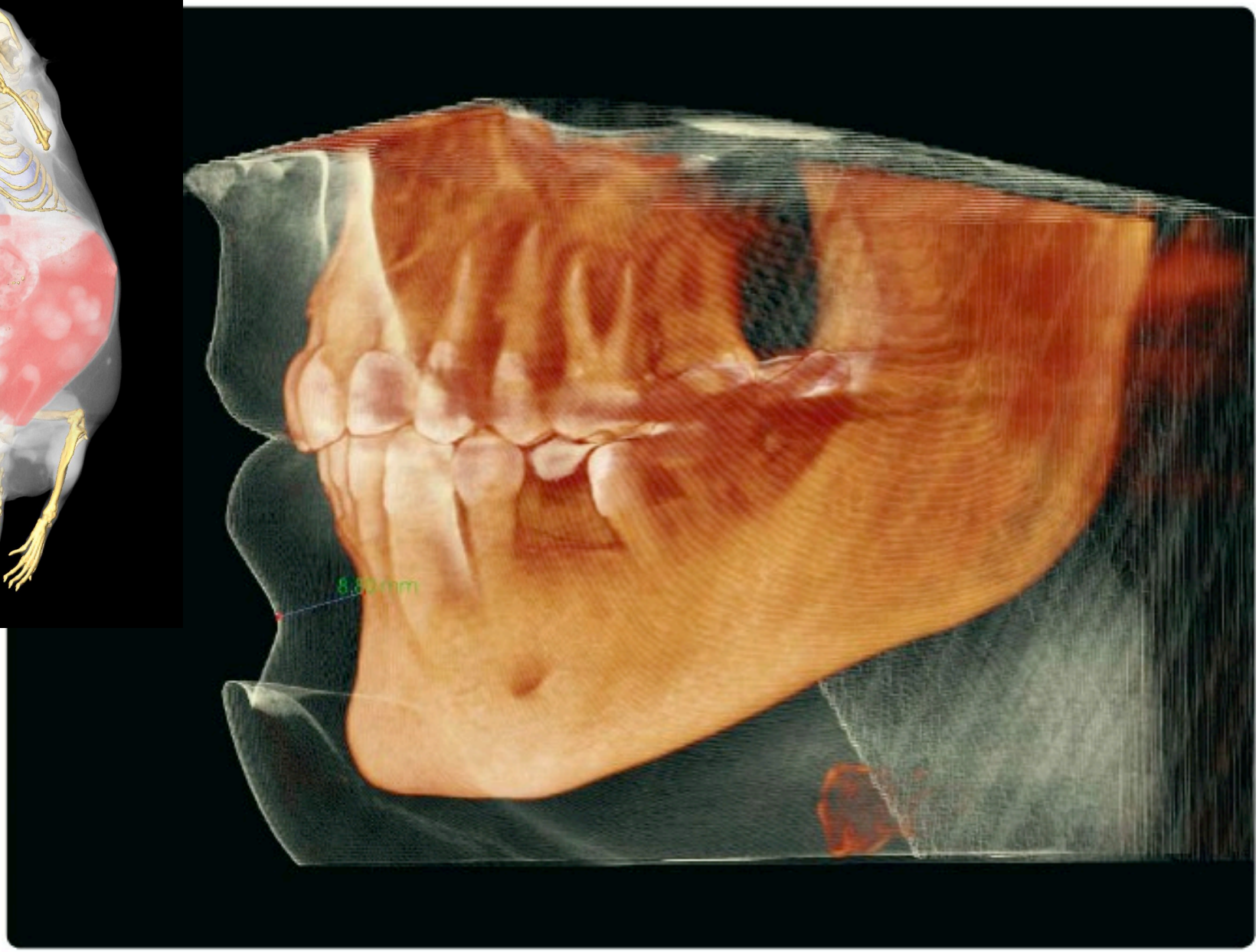
- Confounding circumstances
  - Compute is far away, expensive, batch
  - Storage is distant as well
    - *Datasets are very large*
    - *Disk speed and network bandwidth are constraining*
  - Workstation and display pixels are local
    - *And these are limited in capacity*
- Exploring results is challenging

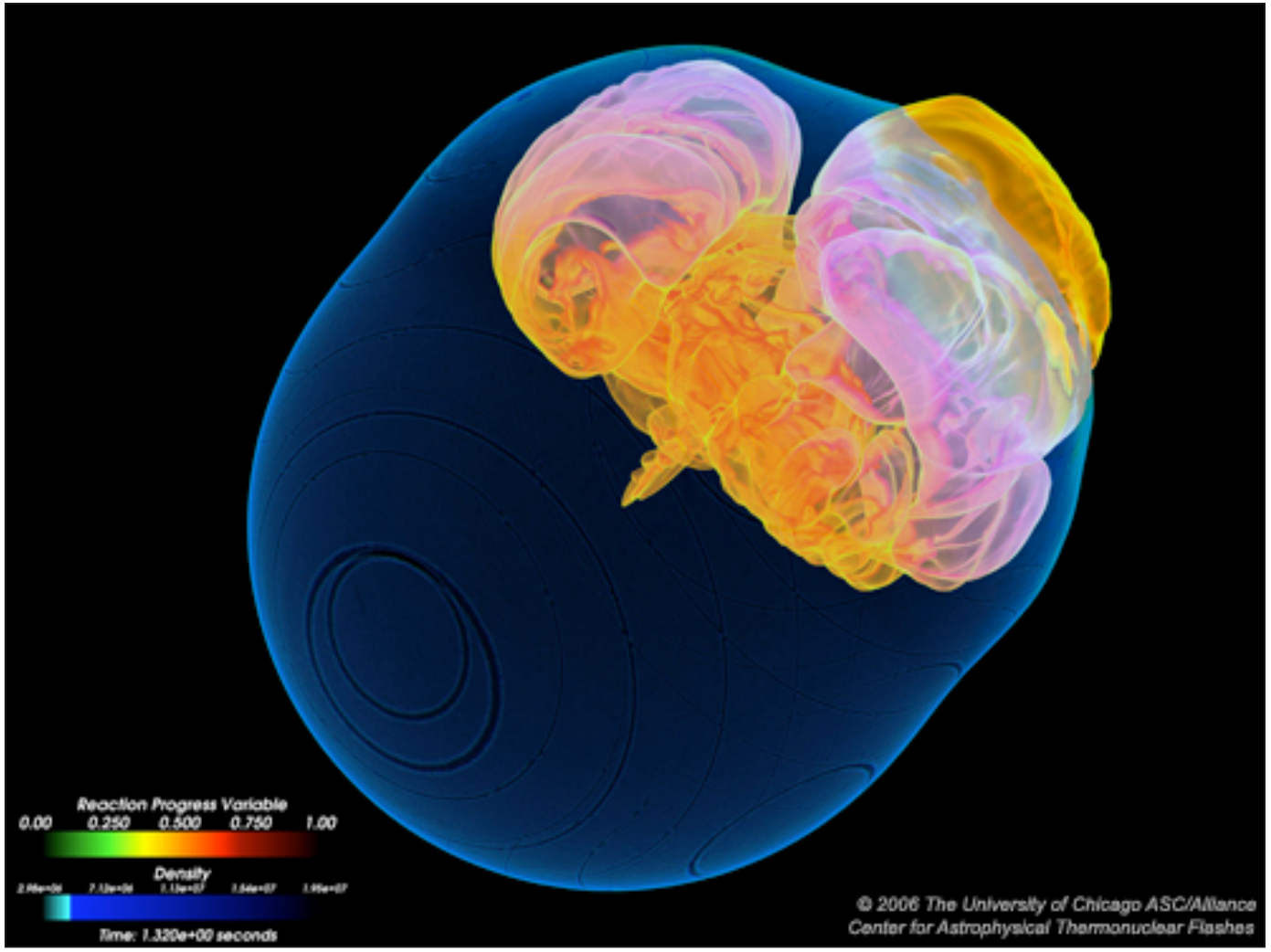
- Data volume example
  - 32K procs each handling 29x29x29 cells
    - *928 x 928 x 928 cells*
  - 751 time steps, 21 variables
    - *30 GB HDF5 file per step, 22 TB total*

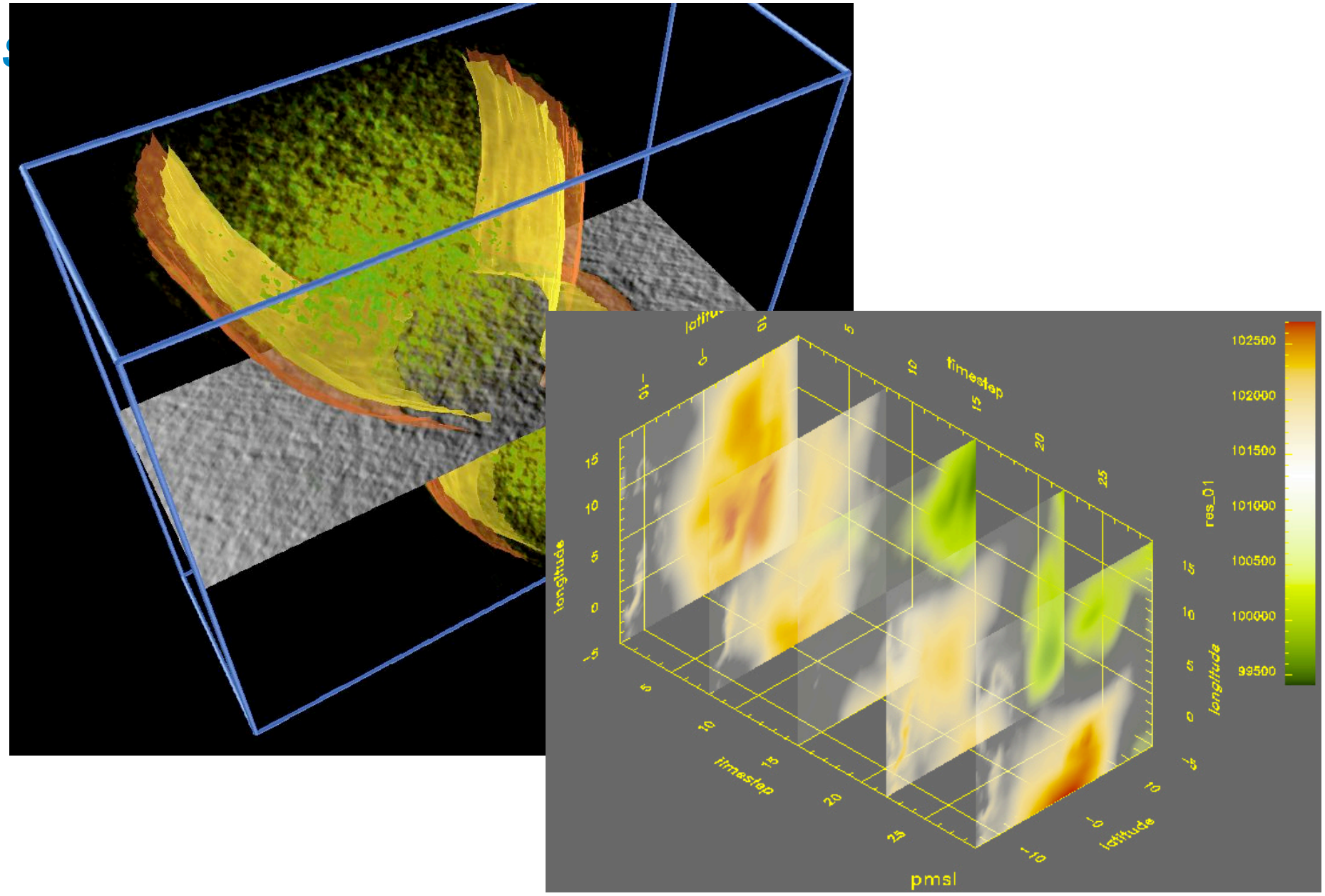
All Sorts of Tools

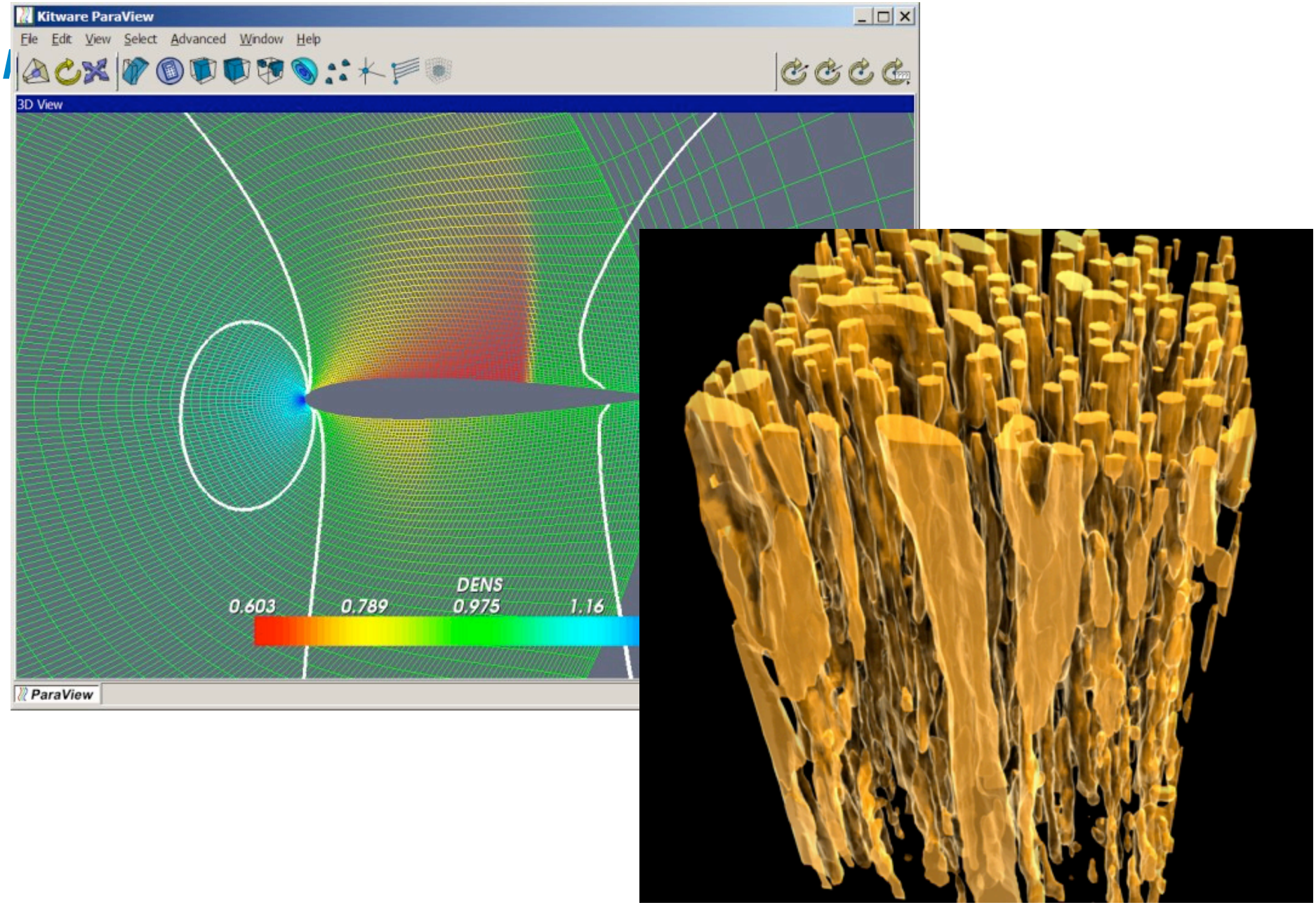
- Visualization Applications
  - VisIt
  - ParaView
  - EnSight
- Domain Specific
  - PyMol, RasMol
- APIs
  - VTK: visualization
  - ITK: segmentation & registration
- GPU performance
  - Scout: GPGPU acceleration
  - vl3: shader-based vol ren
- Analysis Environments
  - Matlab
  - Parallel R (ORNL)
- Utilities
  - GnuPlot
  - ImageMagick
- Visualization Workflow
  - VisTrails

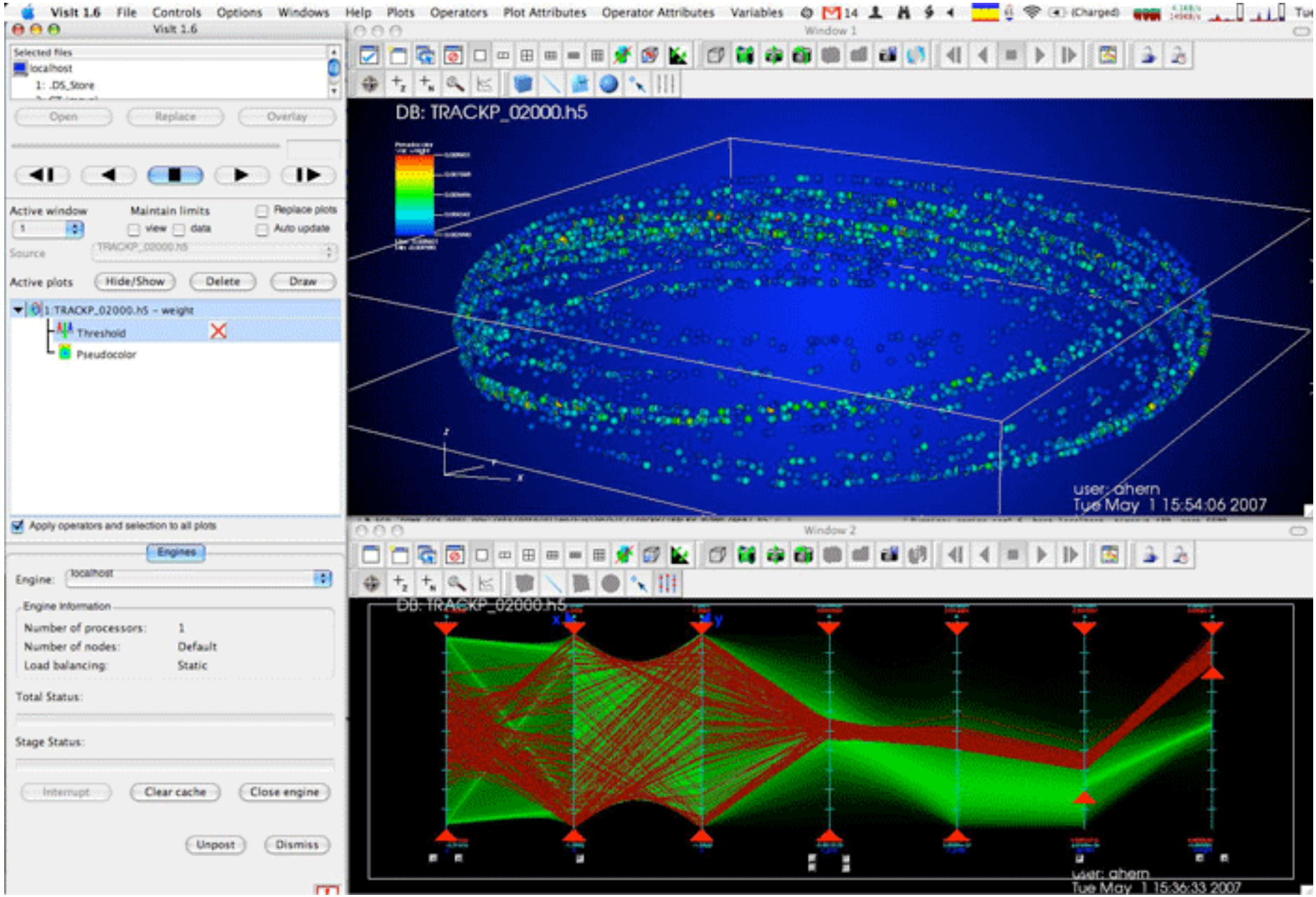




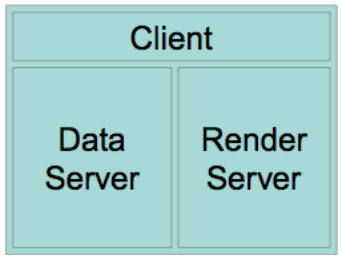

Reaction Progress Variable
0.00
0.250
0.500
0.750
1.00
Density
Time: 1.320e+00 seconds
© 2006 The University of Chicago ASC/Alliance
Center for Astrophysical Thermonuclear Flashes


timestep
longitude
latitude
pmsl
res_01
102500
102000
101500
101000
100500
100000
99500

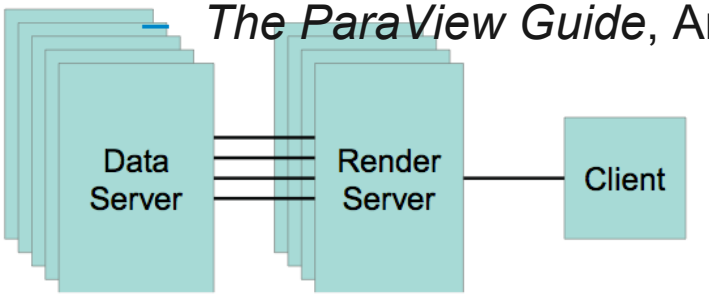

Kitware ParaView
File Edit View Select Advanced Window Help
3D View
DENS
0.603 0.789 0.975 1.16
ParaView

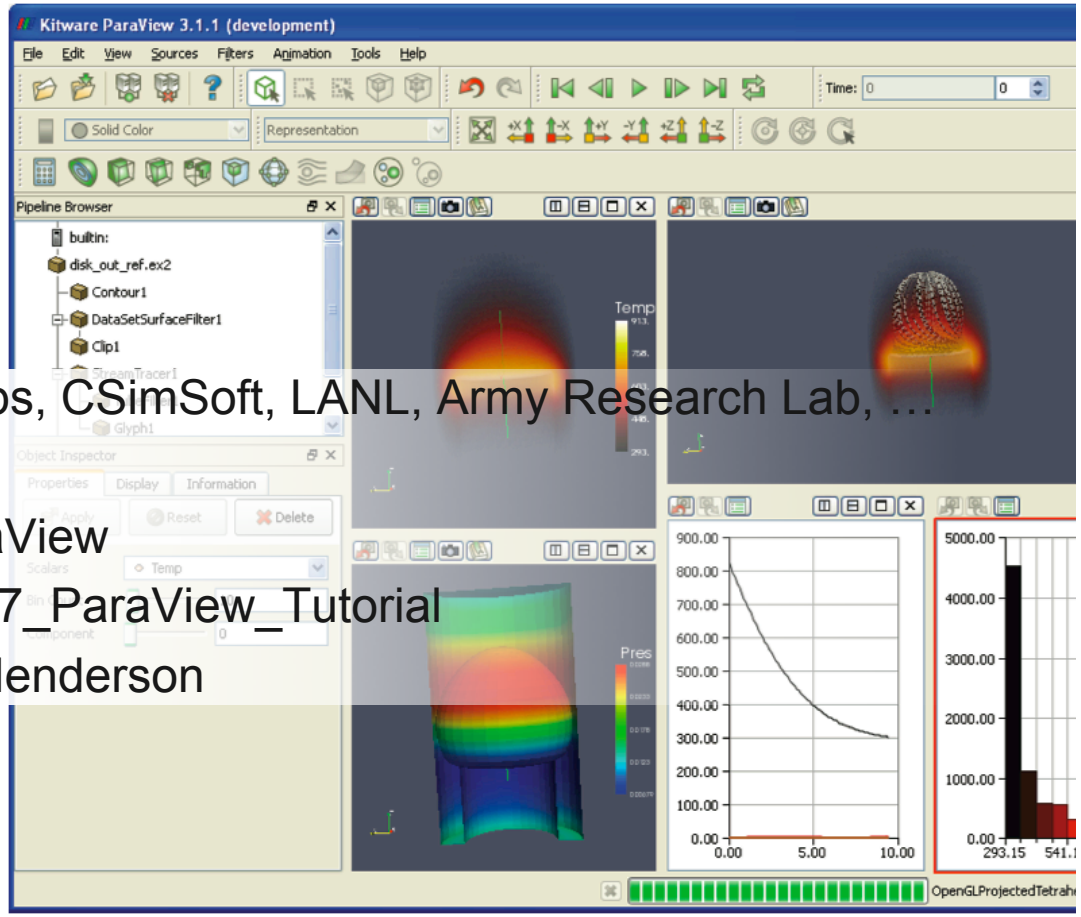

Visit 1.6 File Controls Options Windows Help Plots Operators Plot Attributes Operator Attributes Variables
Visit 1.6
Selected files
localhost
1: .DS_Store
Open
Replace
Overlay
Active window
Maintain limits
view
data
Replace plots
Auto update
Source
TRACKP_02000.h5
Active plots
Hide/Show
Delete
Draw
1:TRACKP_02000.h5 - weight
Threshold
Pseudocolor
Apply operators and selection to all plots
Engines
Engine:
localhost
Engine Information
Number of processors: 1
Number of nodes: Default
Load balancing: Static
Total Status:
Stage Status:
Interrupt
Clear cache
Close engine
Unpost
Dismiss
Window 1
DB: TRACKP_02000.h5
user: ahern
Tue May 1 15:54:06 2007
Window 2
DB: TRACKP_02000.h5
user: ahern
Tue May 1 15:36:33 2007

ParaView Overview



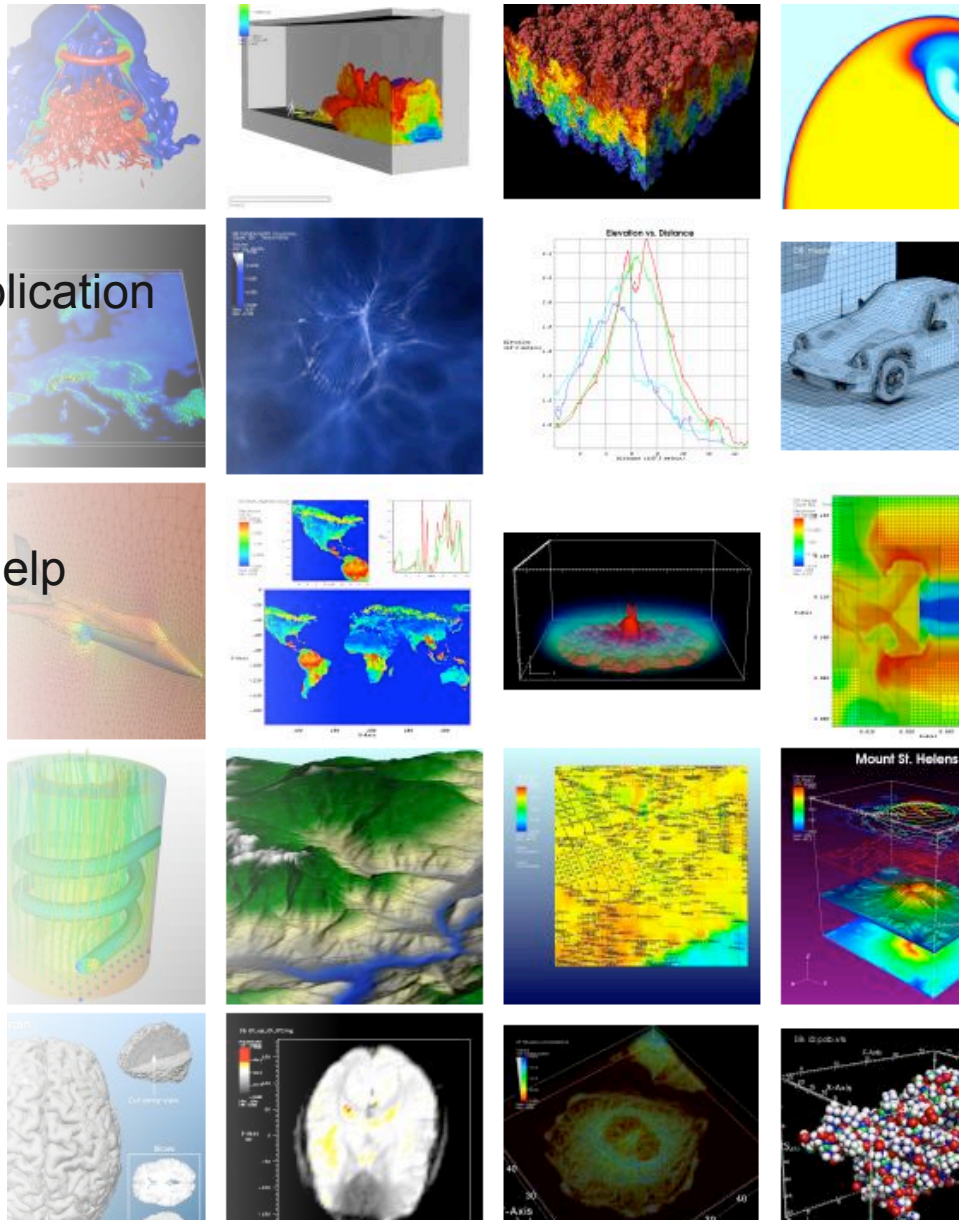
- Parallel Visualization Application
- Open source
- VTK + Tcl
- Python scripting
- Interactive and batch
- About
  - Kitware, Sandia National Labs, CSimSoft, LANL, Army Research Lab, …
  - http://www.paraview.org
  - http://paraview.org/Wiki/ParaView
  - http://paraview.org/Wiki/SC07_ParaView_Tutorial
  - *The ParaView Guide*, Amy Henderson





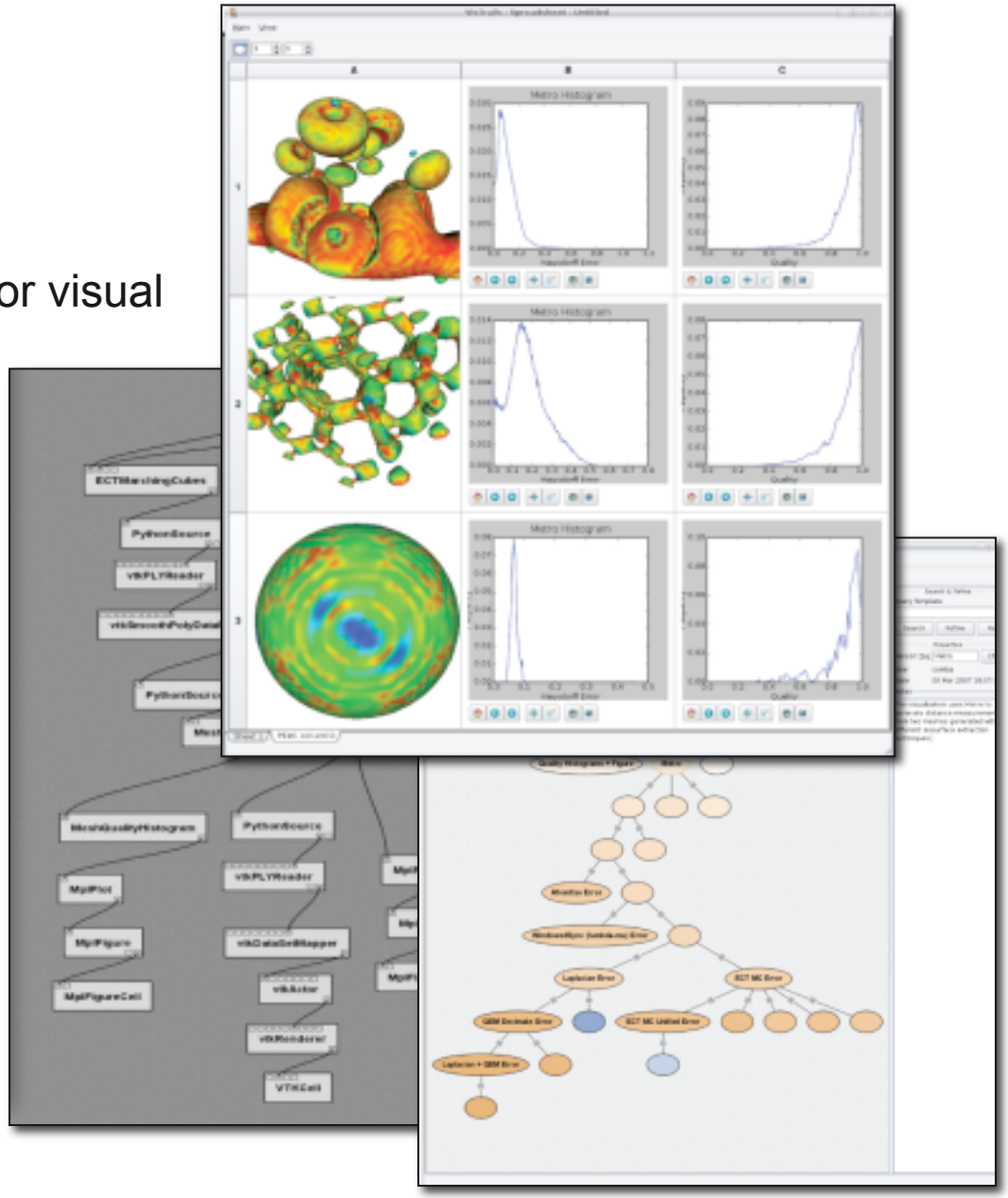
VisIt Overview

- Parallel interactive visualization application
- About
  - DOE ASCI
  - https://www.llnl.gov/visit
  - Manuals, tutorials, application help



VisTrails

- Scientific workflow management for visual data analysis
- Visual programming
- Construct and execute pipelines
  - VTK, ITK, and Matplotlib
- History tree captures provenance
- Visualization spreadsheet
- About
  - http://www.vistrails.org

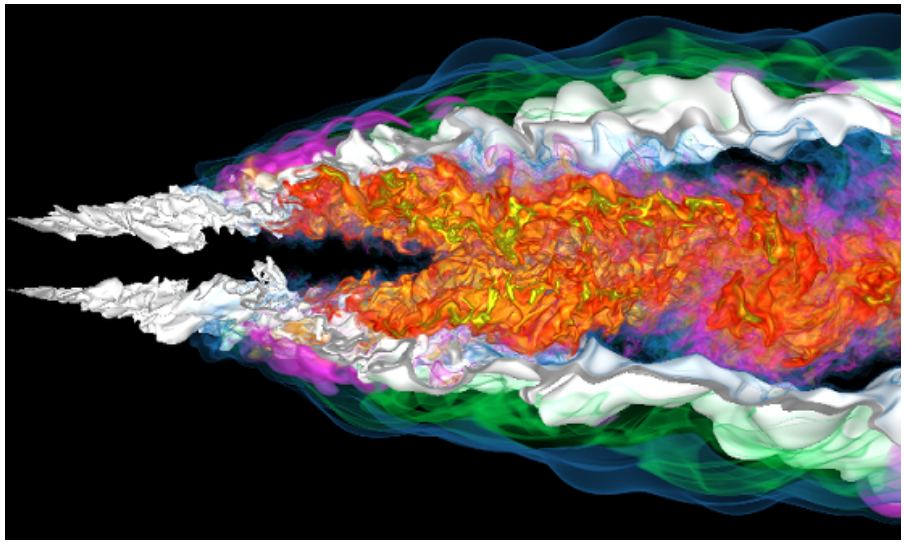


In situ analysis and data reduction

- Incorporate analysis routines into the simulation code
  - operate on data while it is still in memory
- Potential for significant reduction the I/O demands
  - application scientist identifies features of interest
  - compress data of less interest

Here, the feature of interest is the mixture fraction with an iso-value of 0.2 (white surface). Colored regions are a volume rendering of the HO2 variable (data courtesy J. Chen).

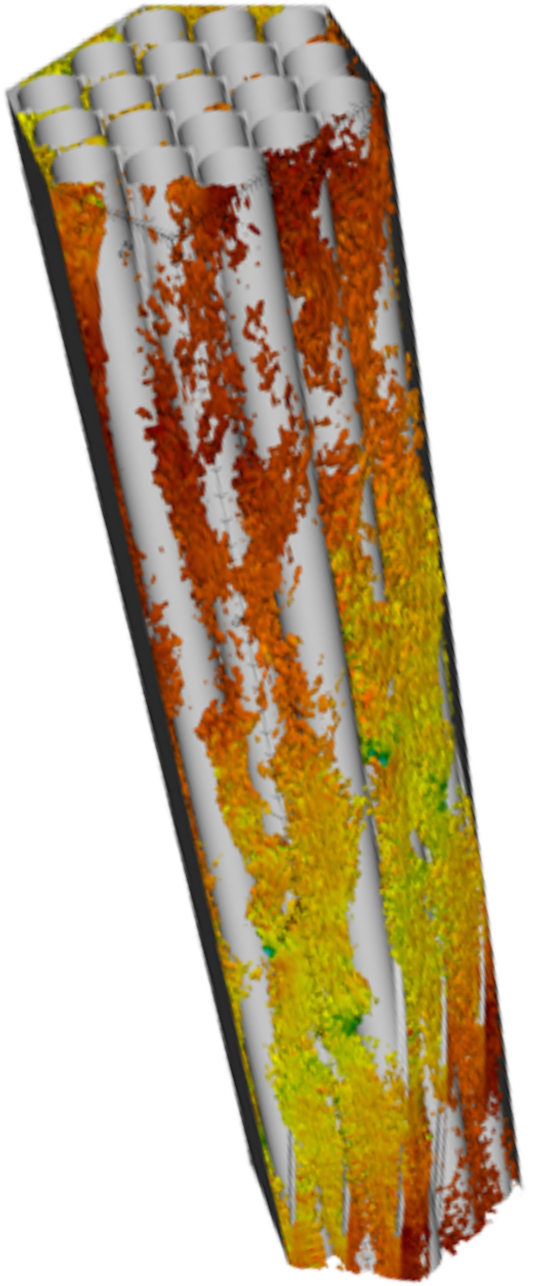
By compressing data more aggressively the further it is from this surface, we can attain a compression ratio of 20-30x while still retaining full fidelity in the vicinity of the surface.



C. Wang, H. Yu, and K.-L. Ma, “Application-driven compression for visualizing large-scale time-varying volume data”, IEEE Computer Graphics and Applications, accepted for publication.

Nuclear Reactor Simulation

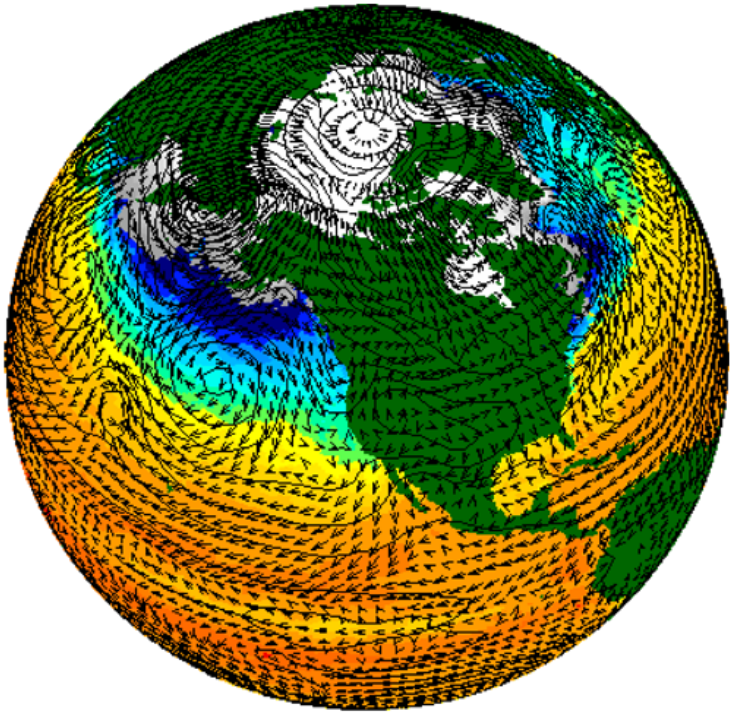
- Preliminary studies
  - 4.5 million elements
  - 7 variables per element
  - 20 K timesteps
  - Total data produced 2.5 TB
- Science runs
  - 3 – 4 runs with 120 million elements
  - Several runs at ½ and ¼ resolution
  - 90 K timesteps
  - Total data produced 900TB – 1.2 PB



Climate Modeling

- Preliminary studies
  - 50-100 with 3 million grid points (1 M atmosphere, 2 M ocean)
  - 100 variables per grid point (30 vectors, 70 scalars)
  - Simulating 5 - 10 years of climate
  - Total data produced 30 -124 TB

- Science runs
  - 50 runs with 6 million grid points
  - Simulating 100 years of climate
  - Total data produced 1.2 PB

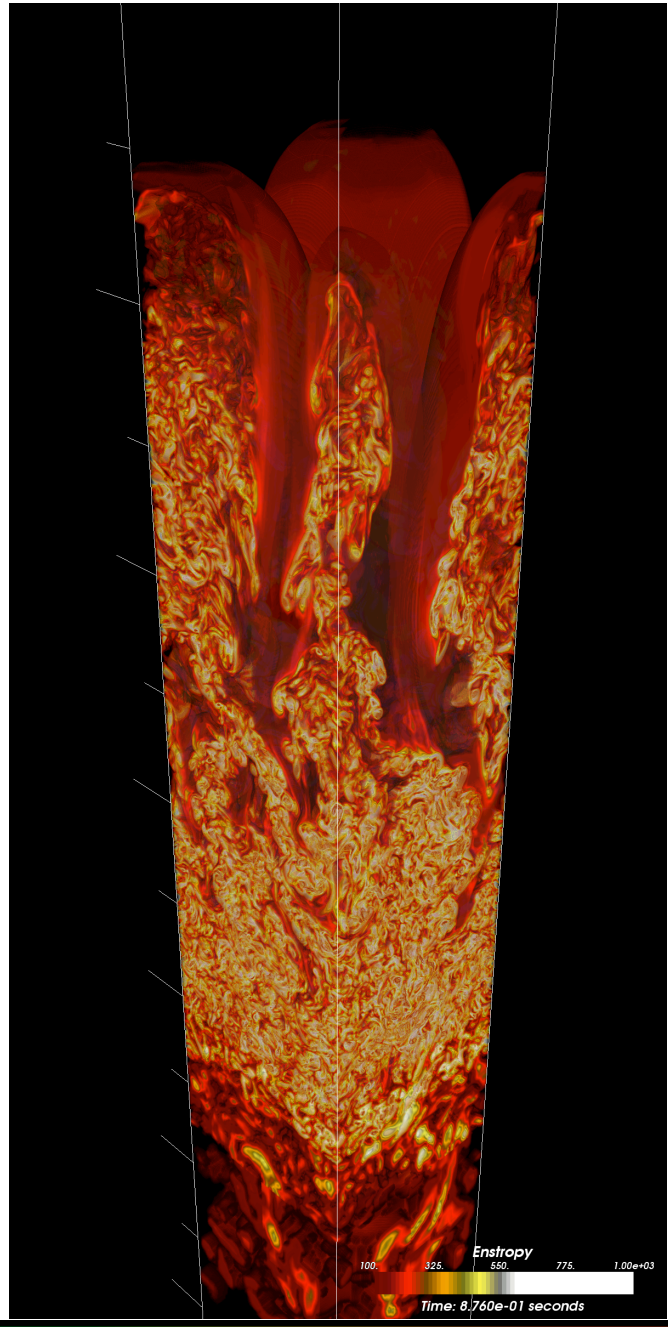


Astrophysics

- Preliminary studies
  - ~80 with 67 M grid points
  - ~5 with 536 M grid points
  - 6 variables (1 vector, 3 scalars)
  - ~1800 time steps
  - Total data produced 78 TB

- Science run*
  - 1024^2 x 4096 grid points
  - 6 variables (1 vector, 3 scalars)
  - ~1800 time steps
  - Total data produced 48 TB

*3-5 times bigger allocation is needed



Analysis/Visualization Questionnaire – Overview

- Do you have preferred computational platform?
- What are your dataset sizes?
  - Do you have checkpoint files and analysis files or is everything together?
- How long do your simulations run? (e.g. wall clock time)
  - Is the result a time series?
  - How many files does that produce?
- Do you do your analysis local or using remote resources?
  - Why?
- How long do you spend on analysis, what is the fraction of compute versus human?

A/V Q – Your I/O footprint

- How much of your simulation time is I/O? (e.g. 10%)
- Do you do data reduction (e.g. convert from double to float during I/O, subsample data, …)?
  - Why?
- What are your transfer rates to disk from simulation? (e.g. bandwidth)
- What are your transfer rates from disk for analysis? (e.g. bandwidth)

A/V Q – Your Goals

- What do you look to get out of analysis process?
- What analysis tools are you currently using?
  - What are the limitations?
- Do you do real-time exploration or batch processing?
  - What is the role of real-time exploration?
    - Batch?
      - Percentage of your analysis time spent in either mode?
- Do you look at images, movies or graphs?
  - What is role of each (e.g. graphs for science, images publications, movies for talks)